Accepted for "Knowledge Based Systems 2001"

Accumulation of Object Representations utilizing Interaction of Robot Action and Perception

Norbert Krüger, Marcus Ackermann, Gerald Sommer

Lehrstuhl für Kognitive Systeme
Institut für Informatik,
Christian–Albrechts–Universität zu Kiel
Preusserstrasse 1-9, 24105 Kiel, Germany
nkr{gs}@ks.informatik.uni-kiel.de

Abstract

We introduce a robotic–vision system which is able to extract object representations autonomously utilizing a tight interaction of visual perception and robotic action within a perception action cycle [12, 24]. Controlled movement of the object grasped by the robot enables us to compute the transformations of entities which are used to represent aspects of objects and to find correspondences of entities within an image sequence.

A general accumulation scheme allows to acquire robust information from imperfect and partly missing information extracted from single frames of an image sequence. Here we use this scheme with a preprocessing stage in which 3D-line segments are extracted from stereo images. However, the accumulation scheme can be used with any kind of preprocessing as long as the entities used to represent objects can be brought to correspondence by certain equivalence relations such as 'rigid body motion'.

We show that an accumulated representation can be applied within a tracking algorithm. The accumulation scheme is an important module of a vision based robot system on which we are currently working. In this system, objects are planned to be represented by different visual and tactile entities. The object representations are going to be learned autonomously. We discuss the accumulation scheme in the context of this project.

1 Introduction

The aim of our research is the design and implementation of an active vision system coupled with a robot arm which is able to recognise and grasp objects with autonomously learned representations. The system shall gain robot control over new objects (i.e., grasp a new object in a scene) by an instinctive and rudimentary behavior pattern and use the control over the object to accumulate a representation of the object and finally apply these representations to robustly track, grasp and recognise the object in complex scenes. Here we describe one module of such a system which can be used to extract object representations.

Model based vision systems usually apply manually designed object representations (see e.g., [27] or [15]). These methods work well but commonly have drawbacks with the need of manual intervention for creating object representations and the

fine–tuning of these representations. Here we demonstrate an autonomous extraction of object representations making use of a tight interaction of perception and action: Accumulation of information takes place within a perception–action–cycle [12, 24]. As a challenging perspective we aim at a coupled robotic–vision system which is not equipped with manually designed object representations but the object to be manipulated is given to the robot and a representation is accumulated autonomously (see figure 1).

Feature extraction faces the problem that semantic information extracted by artificial systems from a single image or stereo images even under optimal conditions is necessarily imperfect. For instance, although there exist a large amount of edge detectors none of them is comparable to human performance. One important reason for the extremely good performance of humans on these tasks is that the human visual system applies *constraints* to interpret a certain scene or situation [7, 13]. A situation never stands for itself but is embedded in a time continuum [8]. Therefore an important constraint is the utilization of the coherence of objects during a rigid body motion which allows to accumulate information over time.

In this paper we suggest to accumulate object representations from image sequences by using the equivalence relation 'rigid body motion'. We account for the vagueness of semantic information extracted from single images by assigning confidences to this information and accumulating this information over an image sequence of a moving object. Although the information extracted from single images contains errors due to, e.g., changes of illumination or noise, (see the representations on the left hand side of figure 1) a more stable representation can be achieved by combining information from different images (see right hand side of figure 1). Because the object can change its position and orientation — and this change might be wanted because another view of the object gives new information which might not be extractable from former ones — we face the correspondence problem: Correspondences between entities describing the object in different images (or 3D interpretations extracted from stereo images) are not known.

Here the correspondence problem is solved within a behavior based paradigm [2, 23]: The parameters of motion are known since the robot manipulates the object and the transformations of entities can be compensated for each frame of the sequence to achieve correspondences. Knowing the correspondences, an algorithm can be applied to update and improve the object representation iteratively. This accumulation algorithm is an extension of an algorithm introduced in [13, 18] which has only dealt with 2D representation and translational motion.

The paper is organized as following. In section 2 we introduce the accumulation scheme and its application to the entity '3D-local line segment'. In section 3 we give a short description of other existing modules of our vision based robot system. Finally, in section 4 we point out to future research.

2 Extraction of Object Representations from Image Sequences

Our accumulation algorithm can be defined independently of the entities used to represent objects. The algorithm also is independent of the concrete equivalence relation or transformation used to define correspondences. It only requires an object representation by certain entities for which a metric is defined and to which certain transformations or equivalence relations (such as rigid body motion) can be applied. The object establishes itself as an invariant under the equivalence relation, i.e., as an equivalence class. The algorithm in its general form is defined in subsection 2.1. In this paper for the representation of objects we use local three dimensional line

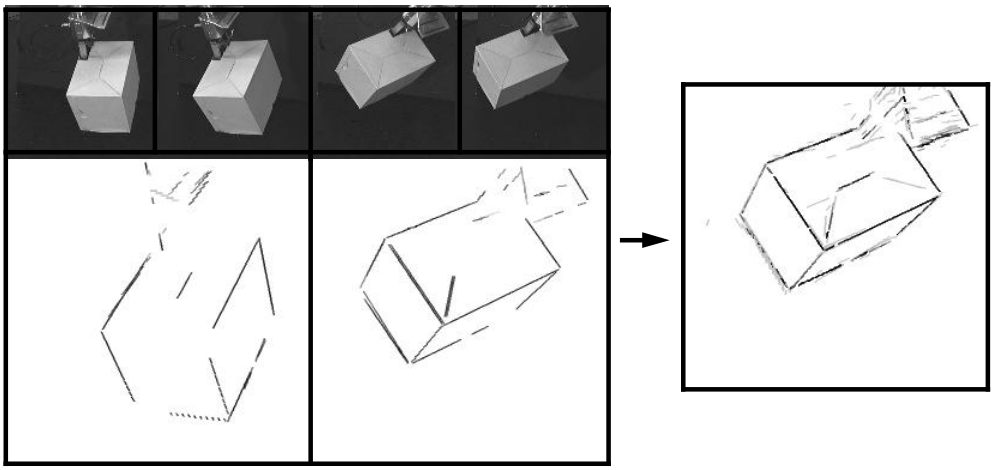

Figure 1: **left)** top: left and right image of an object in two frames. bottom: the projected 3D representation extracted from the stereo images. **right)** Projected 3D Representation accumulated over a set of stereo images. The system's confidence for the presence of line segments is represented as grey value (dark values represent high confidences).

---

segments only. The extension of the system to other kind of object descriptors such as texture or color flow is part of our current research.

The concrete realization of the accumulation scheme can be divided into two parts, preprocessing (section 2.2.1) and accumulation (section 2.2.2). The algorithm is applied to a stereo image sequence in which the object grasped by the robot is shown to the system in various positions and orientations (see figure 1). A representation is accumulated over the stereo image sequence (see figure 1 right). Although the representations extracted from each of the stereo image pairs shows missing line segments (left) the accumulated representation is more complete (right). Here we give a condensed description of the algorithm, for further details see [1].

2.1 The Accumulation Scheme

Let $e \in E$ be an entity used to describe objects (for instance a 2D–line segment, a structure tensor [11] extracted from an image, 3D–line segments extracted from a stereo image pair or any other kind of object descriptor) and $d(e, e')$ be a distance measure on the space of entities E. Furthermore, let T be a transformation or equivalence relation, for instance a rigid body motion or the projection of a rigid body motion. If e^i is an entity extracted from frame i of a sequence of events then $T^{i,i+1}(e^i)$ is the transformation $T^{i,i+1}$ from the i–th to the $i+1$–the frame applied to e^i.

Let e^{i+1} be an entity extracted from the $(i+1)$–th frame of the sequence. We say that e^i and e^{i+1} are likely to correspond to each other if $d(T(e^i), e^{i+1})$ is small. Often it might not be possible to find an exact correspondence with $d(T(e^i), e^{i+1}) = 0$. For example, if we want to compare local image patches in two images knowing the exact projective transformation corresponding to the rigid body motion of an object from the first to the second frame, the corresponding image patches can not be expected to be exactly equal because of factors such as noise during the image acquisition, changing illumination, non–Lambertian surfaces or discretization errors. The problem may even become more severe when we extract more complex entities such as 3D or 2D line segments or 3D–surface patches. Therefore it is advantageous to formalize a confidence of correspondence by using a metric.

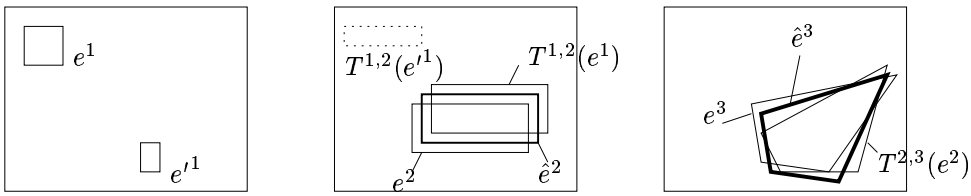

Figure 2: The accumulation scheme. The entity e^1 (here represented as a square) is transformed to $T^{1,2}(e^1)$. Note that without this transformation it is barely impossible to find a correspondence between the entities e^1 and e^2 because the entities show significant differences in appearance and position. Here a correspondence between $T^{1,2}(e^1)$ and e^2 is found because a similar square can be found close to $T^{1,2}(e^1)$ and both entities are merged to the entity $\hat{e}^2$. The confidence assigned to $\hat{e}^2$ is set to a higher value than the confidence assigned to e^1 indicated by the width of the lines of the square. In contrast, the confidence assigned to e'^1 is decreased because no correspondence in the second frame is found. The same procedure is then applied for the next frame for which again a correspondence for e^1 has been found while no correspondence for e'^1 could be found. The confidence assigned to e^1 is once again increased while the confidence assigned to e'^1 is one again decreased (the entity has been disappeared). By this scheme information can be accumulated to achieve robust representations.

The accumulation of information can now simply be achieved by the following update rule: If there exists an entity e^{i+1} in the $(i+1)$–th frame for which $d(T(e^i), e^{i+1})$ is small (i.e. a correspondence is likely), then merge $T(e^i)$ and e^{i+1} by some kind of average operator, $\hat{e}^{i+1} = \text{merge}(T(e^i), e^{i+1})$, and set the confidence for $\hat{e}^{i+1}$ to a higher value than the confidence assigned to e^i. If there exists no entity e^{i+1} in the $(i+1)$–th frame for which $d(T(e^i), e^{i+1})$ is small, the confidence for entity e^i to be part of the object is decreased. In Figure 2 a schematic representation of the algorithm is shown for two iterations.

2.2 Application of the Accumulation Scheme to a Representation with 3D–Line Segments

In this section we apply the accumulation scheme introduced above to object representations consisting of local 3D line segments. For these entities the change of the transformation (i.e., $T^{i,i+1}(e)$) can be computed explicitly (for details see [1]).

2.2.1 Extraction of a 3D Representation from Stereo Images

In the preprocessing step a 3D representation of the object grasped by the robot and presented at a certain position and orientation is extracted. The orientation of the object differs in each stereo image pair (figure 1). The object representation consists of local 3D–line segments and is extracted using calibrated cameras and epipolar geometry [3]. First, in each single image lines are extracted using the orientation sensitive Hough transformation [19]. The Hough lines are divided into local line segments according to local information indicating evidence for the existence of a local line segment at a certain pixel position in the image by evaluating gradient information. In our implementation the entity 'local line segment' can only be extracted when there is local support (a high magnitude of the gradient) *and* global support (the line segment is part of a Hough line). Second, correspondences of line segments in the two stereo images are found. The epipolar constraint is used to reduce the search problem to a one-dimensional problem. On the epipolar line cor-

responding to a certain line segment the best match is defined as the corresponding entity. For finding the best match a similarity combining gray level information (by evaluating the correlation of image patches) and semantic information (evaluating the differences in the orientation of the found line segments) are used.[1]

In most cases the correspondence of 2D line segments defines a 3D line segment. In some cases, when the 2D line segments are close to a 'critical plane' [3, 9] the correspondences do not uniquely define a 3D line segment and a 3D representation of parts of the object can not be extracted. Note that by moving the object, 3D–line segments which can not be extracted in one frame (because they are too close to the critical plane) move out of the critical plane so that they can be part of the final representation. Here the haptic control of the object allows the creation of situations in which critical features can be extracted.

The representation extracted from a single stereo image pair usually is not perfect (see figure 1), there are many missing parts (because of the critical plane, correspondences not found, not detected Hough lines or not extracted 2D line segments in one of the two stereo images) and some 'wrong' line segments (because of wrong correspondences or wrong 2D line segments extracted during preprocessing). Here we face the problem that semantic information can not be extracted with sufficient accuracy from single or stereo images which is also one of the the reasons for the need of manually designed object representations in many artificial systems.

To achieve a suitable representation autonomously and to overcome the need of manual intervention, we accumulate evidence over a self generated stereo image sequence within a perception–action cycle as described in the next subsection.

2.2.2 Accumulation of Object Representations in Stereo Image Sequences

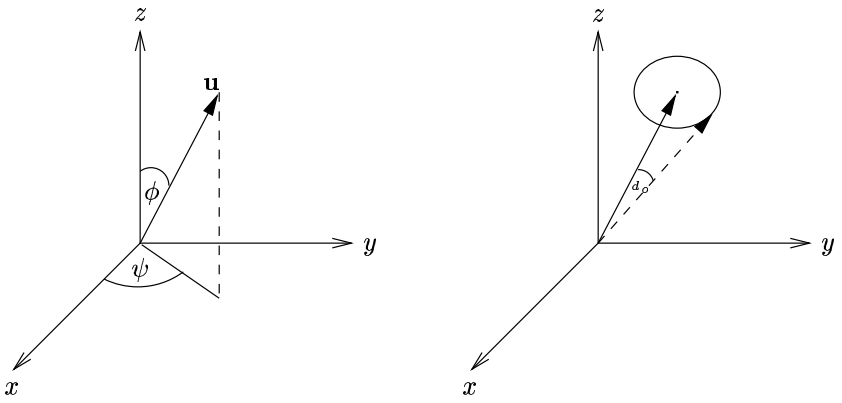

Figure 3: **Left:** Spherical coordinates. **Right:** Difference in orientation d_o

The object representation computed from the first stereo image pair consists of a list $\mathcal{L}$ of 3D line segments $l = (\mathbf{p}, \mathbf{u})$, i.e., a line segment is described by the position of its centre $\mathbf{p} = (x, y, z)$ and by the unit vector $\mathbf{u} = (\phi, \psi)$ indicating the orientation of the line segment (see figure 3 left). For these entities a metric $d(l, l')$ can be defined which gives low values for similar line segments and high values for dissimilar ones.

Definition of the Metric: Here, we define a distance measure $d(l, l')$ between two 3D–line segments *l,l'*. We evaluate the orientation difference d_o and spatial

[1] This kind of preprocessing faces some problems which we currently overcome by use of a filter introduced in [6] instead of the Hough transform. We discuss this issue in detail in section 4.

difference d_p of the line segments. Given $l = (\mathbf{p}, \mathbf{u})$ and $l' = (\mathbf{p}', \mathbf{u}')$. The orientation difference is simply defined as

$$d_o(l, l') := \arccos(\mathbf{u} \cdot \mathbf{u}'),$$

i.e. as the angle between $\mathbf{u}$ and $\mathbf{u}'$ (see figure 3 right).

For the distance measure d_p we have, because of the aperture problem (see e.g. [16]), also to take the orientation of a line segment into account: The translation of a line segment along the axis spanned by $\mathbf{u}$ should not increase the distance between two line segments as long as it is less then half of the length of the line segment. In the following we define an ellipsoid unit sphere, i.e. we allow in the $\mathbf{u}$ direction a larger translation than orthogonal to $\mathbf{u}$ (figure 4 (left) shows the 2D projection of the ellipsoid).

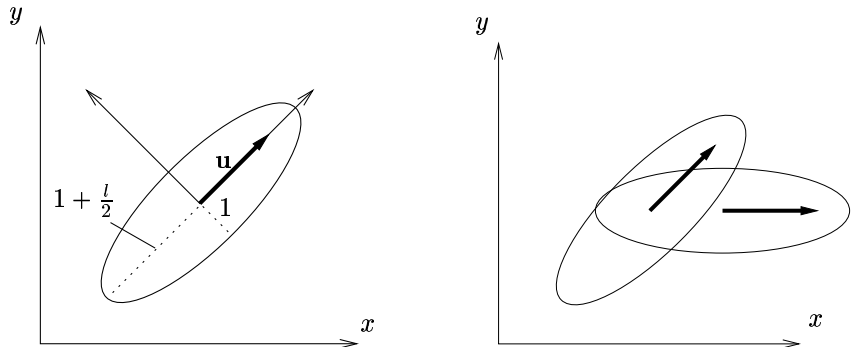

Figure 4: **Left:** Projection of ellipsoid unit sphere. **Right:** Symmetry does not hold for f_p: The midpoint of the diagonal elliptical area is within the horizontal elliptical area but not the other way round.

To compare $\mathbf{p}$ and $\mathbf{p}'$, we need the coordinates $\hat{\mathbf{p}}'$ of $\mathbf{p}'$ in the coordinate system spanned by $\mathbf{p}$ and $\mathbf{u}$ ($\mathbf{p}$ be the origin $(0,0,0)$ and $\mathbf{u} = (\pi/2, 0)$ the x–axis).

We define a distance measure between l and l' (taking into account an elliptical deformation) by

$$f_p(l, l') := \sqrt{\left(\frac{1}{1 + \frac{l}{2}}\hat{x}'\right)^2 + \hat{y}'^2 + \hat{z}'^2}.$$

Since this measure is not symmetric (see figure 4 right) we define

$$d_p(l, l') := \min\{f_p(l, l'), f_p(l', l)\}$$

as the final metric. Now we are able to say that line segment l and l' do correspond to each other when $d_o(l, l')$ and $d_p(l, l')$ are smaller then certain thresholds s_o and s_p.

Accumulation: A rigid body movement M of the robot can be described by six parameters $\vec{\beta} \in I\!R^6$, three describing translation and the others describing rotation. Let $M^{\vec{\beta}}(\mathcal{L})$ be the list of local line segments $\mathcal{L}$ representing the object moved by $M^{\vec{\beta}}$. Let $\mathcal{L}'$ be the list of local line segments extracted from a new stereo image pair. In this image pair the object is shown after a movement whose parameters $\vec{\beta}$ are known. For our algorithm the correspondences between the representations $\mathcal{L}'$ and $\mathcal{L}$ can easily be achieved by applying the rigid body motion $M^{\vec{\beta}}$ to the stored representation $\mathcal{L}$: $M^{\vec{\beta}}(\mathcal{L}) \approx \mathcal{L}'$ and comparison of the line segments by applying

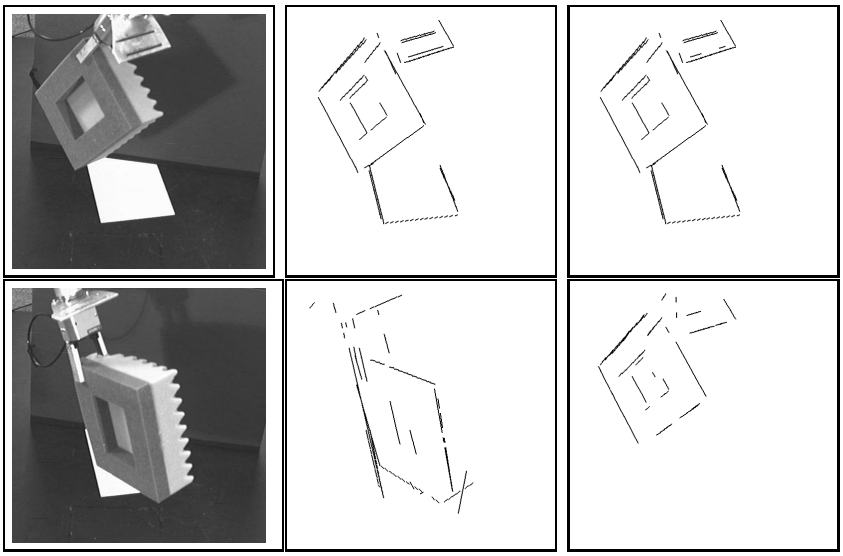

Figure 5: Accumulation of an object representation (first iteration in first row and fifth iteration in second row). Left: one of the stereo images. Middle: Representation extracted from one stereo image pair. Right: Accumulated representation. Line segments corresponding to the background vanish after a few iterations.

---

the above defined metric. The rigid body motion is known in this context, since the robot has physical control over the object.

After achieving correspondences the two representations $M^{\vec{\beta}}(\mathcal{L})$ and $\mathcal{L}'$ can be merged by the accumulation scheme defined above: For each line segment l_j in $M^{\vec{\beta}}(\mathcal{L})$ we search for a line segment l'_k in $\mathcal{L}'$ which is close to l_j according to our metric d. If such a corresponding line segment has been found, a value c_j indicating the confidence of the system that l_j is part of the object is increased, otherwise it is decreased. Line segments in $\mathcal{L}'$ to which no correspondences in $M^{\vec{\beta}}(\mathcal{L})$ do exist are included in the accumulated representation with only low confidences. After a couple of iterations with different views of the object the accumulated representation becomes more and more stable (see figure 1 right). It is even possible to segment objects from the background: Since the background is fixed and not changing according to the equivalence relation rigid body motion, line segments corresponding to the background do vanish after a few iterations (see figure 5) and only line segments corresponding to the object and gripper remain.

Note that in this scheme, an entity (here, a 3D line segment) is regarded to be existent only if it has accumulated confidence over time, or more precise, it is understood as an invariant entity in the time–space continuum under the equivalence relation 'rigid body motion'. Therefore an interpretation (as a 3D–line segment) is grounded in its change under the controlled movement of the object: the entity '3D–line segment' establishes itself only if it has been reconfirmed within the perception–action cycle. Our ansatz is therefore related to the so called symbol grounding problem (see [10]), i.e., to the problem to assign meaning to abstract entities. Here 'meaning' can be interpreted as an observable and foreseeable change under a self–performed motion.

3 The Accumulation Scheme as a Module in a Vision Based Robot System

The introduced module is part of our effort to design and implement a vision based robot which is able to recognise and grasp objects with autonomously learned representations. The system shall gain robot control over new objects (i.e., grasp a new object in a scene) by an instinctive and rudimentary behavior pattern and use the control over the object to accumulate a representation of the object and finally apply these representations to robustly track, grasp and recognise the object in a complex scene.

The design of our system is guided by a behavior based paradigm (see [2, 23]) in a dual sense. Firstly, to perform a certain action we may only need to extract a minimum of information (e.g., to fixate and zoom we do not need exact shape information in our system). This is *perception for action*. Secondly, by active intervention we can make tasks easier for perception (e.g., in our system fixating and zooming potentially facilitates grasping or, as another example, robot control over the object helps for the extraction of object representations). This is *action for perception*. Usually both aspects — perception for action and action for perception — occur together in a so called *perception–action cycle* (PAC) [12, 24], i.e., perception and action support each other and depend on each other permanently.

We think that a complex vision–based system can not start to learn without some kind of prior knowledge [7, 13]. It can neither be a fully predetermined system, because the world within it operates is too complex that algorithms which solve difficult tasks could be formalised explicitly. Nor can it be a fully undetermined structure because the space of possible algorithms to be explored is much too large. Therefore, a certain amount of a priori knowledge has to be built in a complex vision system to guide learning. We think that an important part of this knowledge are basic competences (as our accumulation scheme), necessary to start a bootstrapping process in which more complex competences can be established.

A second module of the system is a visual and (potentially) haptic attention mechanism described more precisely in [14]). In this module the system directs its attention to new objects and manipulates the active components (i.e., cameras and grasper) such that a situation is achieved in which grasping becomes easier: grasper and object appear in the centre of a zoomed stereo image pair. In this situation grasping of the object can be performed using only relative positions between grasper and object. The high resolution allows to accurately extract 3D–information about the relative position and orientation of grasper and object by stereo. Our attention module is combined by a number of more primitive competences such as detection of a new object, fixation, a simple recognition scheme of an object under controlled conditions after fixation, movement of the grasper, zoom etc. Note that our attention mechanism is planned not to be only vision–based. We are currently redeveloping a haptic sensor [22] which allows to explore an object haptically. Therefore, our attention mechanism potentially focuses visual *and* haptic attention to the new object. The attention mechanism is to a wide degree predetermined but also contains adaptable components: The grasper is permantly tracked by the system. The information of motor commands and positioning of the grasper in the image allow a self–calibration during the perception–action cycle.

As a third module we apply a novel 2D-3D pose estimation algorithm [26] to the tracking problem (described more precisely in [20]). This pose estimation algorithm shows some interesting characteristics which makes it especially useful for this purpose. Beside features such as stability in the presence of noise and online–capabilities [21, 28] its main advantage in the tracking context is that it can unify different kinds of correspondences within the algebraic framework of geometric alge-

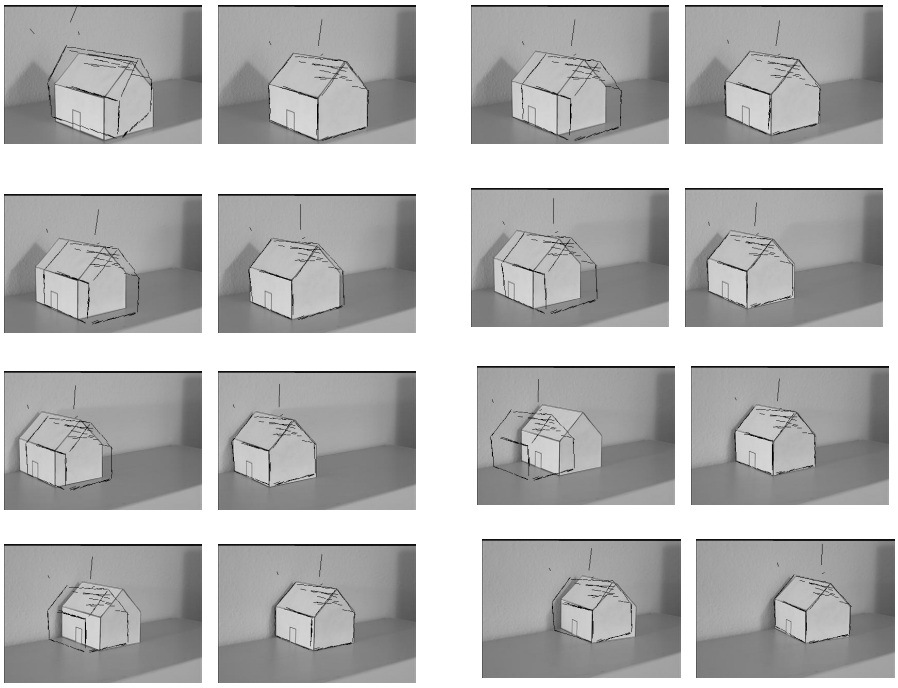

Figure 6: Tracking with an accumulated object model. In each of the image pairs the pose of the object before (left) and after pose estimation (right) is shown.

bra (for a discussion of the application of geometric algebra in computer vision see [25]). We applied the tracking algorithm with an accumulated object representation consisting of local 3D line segments. Figure 6 shows the succesful tracking with such a learned representation. An interesting extension of the accumulation scheme could be the combination of tracking and accumulation, or, more precisely, instead of using the parameters of the rigid body motion from the motor commands given to the robot we can use the parameters estimated by our pose estimation algorithm. In this way the competences 'tracking' and 'accumulation' can support each other.

Here we like to point out a further problem of our object representation: An object is coded by a large number of satistically dependent *local* entities. The use of local features is sensible for accumulation because of two reasons. First, an object usually is bounded in space and second, for local entities it is easier to define correspondences than for more complex features which are also more difficult to extract from images. Nevertheless, for matching it would be advantageous if these entities become connected by some kind of grouping process to achieve a representation with a smaller set of more complex features to speed up matching. The formalization of such perceptual grouping processes (for an overview see, e.g., [17]) is part of our research.

In its current state, only these three modules are implemented. However, in section 4 we give a short overview about our current and future research aiming at a more complete system.

4 Conclusion and Outlook

We showed that our algorithm is able to accumulate autonomously representations utilizing self–controlled movements within a perception–action cycle. For the future a robot systems equipped with the ability to extract efficient object representations

in a normal environment promises more flexible applications of robot vision systems. Instead of being equipped with manually defined representations the robot may use its own ability as a basis for manipulation and recognition. A first stage of our algorithm could be a behavior which allows to achieve haptic control over new objects and which positions the robot arm and the camera such that the accumulation process can start. The bridge between the visual-haptic attention mechanism and the accumulation module, i.e. grasping after fixation and zooming, still has to be done. However, for such a grasp the attention mechanism gives a good starting point, because we have only to operate with relative positions and since we gained high resolution of the important aspects of the scene by active control of the camera. Furthermore, we intend also to use haptic information for performing grasping.

As already pointed out in section 2.2.1 our 3D reconstruction scheme faces some problems: First, for small line segments Hough lines often are not found within the Hough array. Second, it is only the geometric interpretation of the 3D–line segment which is included in the object representation. To overcome these problems we aim to use the structure multi-vector introduced in [6]. This operator, also derived within the framework of geometric algebra, is an isotropic extension of the analytic signal to two dimensions. It has some interesting properties in the context here. First, it is a *local* filter in contrast to the global Hough transformation. Second, the structure multi–vector performs a local decomposition of the signal into energetic, structural and geometric information (split of identity, see [5]). The separated information about geometry and structure can be used for matching on the epipolar line (for details, see [4]). Third, it allows to include both, geometric and structural (i.e., appearance based) information within the object representation. We think, that also this kind of information can be accumulated by our scheme because of its generic structure. We think it is advantageous to use different kind of information of the objects, 3D-based or/and appearence based, depending on the actual task. It is likely that for,e.g., recognition tasks the appearance based aspects of the objects are more significant than the 3D-aspects. However, for grasping or navigation the 3D-aspects might be more important. The aim of our approach is a representation which includes both aspects.

A further important problem is the accumulation of the complete 3D-structure of the object. Up to know only one aspect of an object can be accumulated because correspondences are needed which are not granted when occlusion does occur. That means, when the robot rotates the object by a larger angle it is likely that new edges occur in the stereo images and other edges disappear. An important extension of our algorithm would be the full 3D representation of objects, i.e. the algorithm should extend the representation when, due to occlusion, a new aspect is presented.

We aim at a multi-modal representation of objects containing visual entities such as contour information, texture or colour embedded within a unified framework. This representation shall also contain haptic information (such as roughness or size). Both kinds of descriptions — visual and haptic — can support each other. For example the visual estimated size of an object can be compared with the distance of grasper jaws after grasping.

The design of a vision–based robot system in which basic competences (such as introduced here) interact with each other to derive more complex behavior patterns is a challenging and demanding perspective. It desires the integration of different disciplines such as robotics, computer vision, signal processing and statistical learning as well as the integration of software developed by different people. Finally, the success of such a system should be measured empirically.

Acknowledgment

We thank Josef Pauli and Laurenz Wiskott for fruitful discussions and Christian Perwass for his feedback concerning the English. Our special thanks for Bodo Rosenhahn and Torge Rabsch for their tracking simulations with our accumulated representations and Yiwen Zhang for the implementation of the Kalman filter used within the pose estimation algorithm. Last but not least, we would like to thank Kord Ehmke, Oliver Granert, Daniel Grest, Marco Hahn, Torge Rabsch, Volker Rölke and Bodo Rosenhahn whose work at the software library KiViGraP was very helpful for our simulations. For technical support we would like to thank Gerd Diesner and Henrik Schmidt.

References

[1] M. Ackermann. Akkumulieren von Objektrepräsentationen im Wahrnehmungs–Handlungs Zyklus. *Christian–Albrechts Universität zu Kiel, Institut für Informatik und Praktische Mathematik (Diplomarbeit)*, 2000.

[2] R.A. Brooks. Intelligence without reason. *International Joint Conference on Artificial Intelligence*, pages 569–595, 1991.

[3] O.D. Faugeras, editor. *Three–Dimensional Computer Vision.* MIT Press, 1993.

[4] Michael Felsberg, Norbert Krüger, Martin Poerksen, and Gerald Sommer. Split of identity applied for 3D-reconstruction with the structure multi-vector. *work in progress.*

[5] Michael Felsberg and Gerald Sommer. The monogenic signal. *Christian Albrechts Universität Kiel, Institut für Informatik und Praktische Mathematik, Technical Report No. 2006*, 2000.

[6] Michael Felsberg and Gerald Sommer. Structure multivector for local analysis of images. *Christian Albrechts Universität Kiel, Institut für Informatik und Praktische Mathematik. Technical Report No. 2001*, 2000.

[7] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1995.

[8] J.J. Gibson. *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin, 1979.

[9] Marco Hahn. Semiglobale Verfahren zur Generierung von Eckpunkthypothesen in 2D und 3D. *Christian–Albrechts Universität zu Kiel, Institut für Informatik und Praktische Mathematik (Diplomarbeit)*, 1999.

[10] S. Harnad. The symbol grounding problem. *Physica*, D(42):335–346, 1990.

[11] B. Jähne, editor. *Digitale Bildverarbeitung.* Springer, 1997.

[12] J.J Koenderink. Wechsler's vision: An essay review of computational vision by Harry Wechsler. *Ecological Psychology*, 4:121—128, 1992.

[13] N. Krüger. *Visual Learning with a priori Constraints.* (PhD Thesis) Shaker Verlag, Germany, 1998.

[14] Norbert Krüger, Daniel Wendorff, and Gerald Sommer. Two models of a vision–based robotic system: Visual haptic attention and accumulation of object representations. *Accepted for Robot Vision 2001.*

[15] A. Lanitis, C.J. Taylor, and T.F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 7:743–756, 1997.

[16] H.A. Mallot. *Sehen und die Verarbeitung visueller Information.* Vieweg, 1998.

[17] Stefan Posch. *Perzeptives Gruppieren und Bildanalyse.* Habilitationsschrift, Universität Bielefeld, Deutscher Universitäts Verlag, 1997.

[18] M. Pötzsch, N. Krüger, and C. von der Malsburg. A procedure for automatic analysis of images and image sequences based on two–dimensional shape primitives. *U.S. Patent Application*, 1999.

[19] J. Princen, J. Illingworth, and J. Kittler. An optimizing line finder using a Hough transform algorithm. *Computer Vision, Graphics, and Image Processing*, 52:57–77, 1990.

[20] B. Rosenhahn, N. Krüger, T. Rabsch, and G. Sommer. Automatic tracking with a novel pose estimation algorithm. *accepted for Robot Vision 2001*, 2001.

[21] B. Rosenhahn, Y. Zhang, and G. Sommer. Performance of constraint based pose estimation algorithms. In G. Sommer, N. Krüger, and Ch. Perwass, editors, *Mustererkennung 2000*, pages 277–285. Springer Verlag, 2000.

[22] Peer Schmidt. Entwicklung und Aufbau von taktiler Sensorik für eine Roboterhand. *Institut für Neuroinformatik Bochum (Internal Report)*, 2000.

[23] G. Sommer. Verhaltensbasierter Entwurf technischer visueller Systeme. *Künstliche Intelligenz*, 3:42–45, 1995.

[24] G. Sommer. Algebraic aspects of designing behaviour based systems. In G. Sommer and J.J. Koenderink, editors, *Algebraic Frames for the Perception and Action Cycle*, pages 1–28. Springer Verlag, 1997.

[25] G. Sommer. The global algebraic frame of the perception–action–cycle. In B. Jähne, H. Haussecker, and P. Geißler, editors, *Handbook of Computer Vision and Apllications*, volume 3, pages 221–264. Acadamic Press, 1999.

[26] G. Sommer, B. Rosenhahn, and Y. Zhang. Pose estimation using geometric constraints. *Christian–Albrechts–Universität zu Kiel, Institut für Informatik und Praktische Mathematik, Technischer Bericht 2003*, 2000.

[27] Alan L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.

[28] Y. Zhang, B. Rosenhahn, and G. Sommer. Extended Kalman filter design for motion estimation by point and line observations. In *Second international workshop, Algebraic Frames for the Perception-Action Cycle, AFPAC 2000, LNCS 1888.* Springer Verlag, 2000.